

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号
特開2018-149629
(P2018-149629A)

(43) 公開日 平成30年9月27日(2018.9.27)

(51) Int. Cl.	F I	テーマコード (参考)
B 2 5 J 13/00 (2006.01)	B 2 5 J 13/00 Z	2 C 1 5 0
B 2 5 J 13/08 (2006.01)	B 2 5 J 13/08 Z	3 C 7 0 7
G 1 0 L 15/10 (2006.01)	G 1 0 L 15/10 5 0 0 N	
G 1 0 L 25/63 (2013.01)	G 1 0 L 25/63	
G 1 0 L 25/15 (2013.01)	G 1 0 L 25/15	

審査請求 未請求 請求項の数 11 O L (全 17 頁) 最終頁に続く

(21) 出願番号	特願2017-47509 (P2017-47509)	(71) 出願人	393031586 株式会社国際電気通信基礎技術研究所 京都府相楽郡精華町光台二丁目2番地2
(22) 出願日	平成29年3月13日 (2017.3.13)	(74) 代理人	100099933 弁理士 清水 敏
(出願人による申告) 平成26-28年度 国立研究開発法人科学技術振興機構 戦略的創造研究推進事業 総括実施型研究E R A T O「石黒共生ヒューマンロボットインタラクションプロジェクト」に係る委託業務、産業技術力強化法第19条の適用を受ける特許出願		(72) 発明者	イシイ カルロス トシノリ 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
		(72) 発明者	港 隆史 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
		Fターム(参考)	2C150 AA05 CA01 DA25 DA27 DF03 DF33 ED42 EF29 3C707 AS36 KS11 KS39 MT14 WA03 WL07 WL15 WL16 WM25

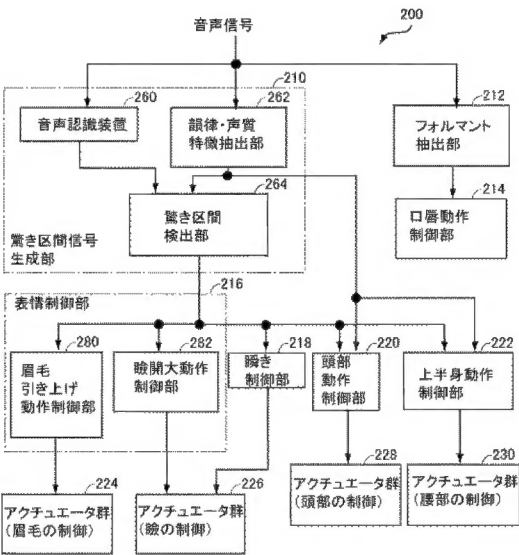
(54) 【発明の名称】 ヒューマノイドロボットの驚き動作生成装置

(57)【要約】

【課題】ヒューマノイドロボットにおいて、驚きに伴う、自然な、人らしい動作を実現する。

【解決手段】驚き動作生成装置200は、驚き区間信号を受信し、開始時刻に、表情を驚き状態に変化させる動作を開始し、終了時刻の後、第1の戻り時間内に表情を中立状態に戻すようロボットを制御する表情制御部216と、開始時刻に頭部を驚き状態に変化させる動作を開始し、終了時刻の後、第2の戻り時間内に、頭部の状態を中立状態に戻すようロボットを制御する頭部動作制御部220と、開始時刻に上半身を驚き状態に変化させる動作を開始し、終了時刻の後、第3の戻り時間内に上半身を中立状態に戻すようロボットを制御する上半身動作制御部222を含む。これらのいずれにおいても、戻り時間が、驚き状態への変化に要する時間よりも長くなるように制御される。

【選択図】図12



【特許請求の範囲】

【請求項1】

所定の時間区間において、驚きに伴う動作を行う様にヒューマノイド型のロボットを制御するための驚き動作生成装置であって、

前記ロボットの顔の表情、頭部の状態、及び上半身の状態は、驚いていない状態である中立状態と、驚いた状態である驚き状態とにそれぞれ制御可能であり、

前記時間区間の少なくとも開始時刻と終了時刻とを特定する驚き区間信号を受信する様に接続され、前記開始時刻に、前記顔の表情を前記驚き状態に変化させる動作を開始し、前記終了時刻の後、第1の戻り時間内に、前記顔の表情を前記中立状態に戻すよう、前記ロボットを制御する表情制御手段と、

10

前記開始時刻に、前記頭部を前記驚き状態に変化させる動作を開始し、前記終了時刻の後、第2の戻り時間内に、前記頭部の状態を前記中立状態に戻すよう、前記ロボットを制御する頭部制御手段と、

前記開始時刻に、前記上半身を前記驚き状態に変化させる動作を開始し、前記終了時刻の後、第3の戻り時間内に、前記上半身を前記中立状態に戻すよう、前記ロボットを制御する上半身制御手段とを含み、

前記表情制御手段、前記頭部制御手段、及び前記上半身制御手段のいずれにおいても、前記第1、第2又は第3の戻り時間が、前記開始時刻に開始した、前記驚き状態への変化に要する時間よりも長くなる様に制御される、驚き動作生成装置。

20

【請求項2】

前記驚き状態は、第1のレベルの驚き状態と、前記第1のレベルより高い第2のレベルの驚き状態とを含み、

前記驚き区間信号に加えて、驚きのレベルを示す驚きレベル信号がさらに前記驚き動作生成装置に与えられ、

前記表情制御手段、前記頭部制御手段、及び前記上半身制御手段の任意の組み合わせにおいて、前記驚きレベル信号に応じて驚き状態を前記第1のレベルに応じた第1の驚き状態と、前記第2のレベルに応じた第2の驚き状態とに区別して前記ロボットを制御する、請求項1に記載の驚き動作生成装置。

【請求項3】

前記中立状態から前記第2の驚き状態への前記ロボットの各部の変位量は、前記中立状態から前記第1の驚き状態への前記ロボットの各部の変位量より大きい、請求項2に記載の驚き動作生成装置。

30

【請求項4】

前記表情制御手段は、前記驚き状態における眉毛の目からの離間量が、前記中立状態における離間量より大きくなる様に眉毛を制御する、請求項1～請求項3のいずれかに記載の驚き動作生成装置。

【請求項5】

前記頭部制御手段は、前記驚き状態における前記頭部が、前記中立状態における位置から前記ロボットにとって後方に移動する様に前記頭部を制御する、請求項1～請求項4のいずれかに記載の驚き動作生成装置。

40

【請求項6】

前記上半身制御手段は、前記驚き状態における前記上半身が、前記中立状態における位置から前記ロボットにとって後方に反る様に前記上半身を制御する、請求項1～請求項5のいずれかに記載の驚き動作生成装置。

【請求項7】

前記表情制御手段、前記頭部制御手段、及び前記上半身制御手段のいずれも、前記中立状態と前記驚き状態との間の変化が滑らかになる様に前記ロボットを制御する、請求項1～請求項6のいずれかに記載の驚き動作生成装置。

【請求項8】

前記顔の表情が、前記驚き状態から前記中立状態に戻ったことに応答して、前記ロボット

50

の目を1回閉じさせた後、前記中立状態に戻す制御を行うことにより前記ロボットに瞬きを行わせるための瞬き制御手段をさらに含む、請求項1～請求項7のいずれかに記載の驚き動作生成装置。

【請求項9】

前記時間区間において、前記ロボットが発声すべき発話音声信号のフォルマントを抽出するフォルマント抽出手段と、

前記フォルマント抽出手段により抽出されたフォルマントに対応して前記ロボットの口唇の開口量を制御するための口唇動作制御手段をさらに含む、請求項1～請求項8のいずれかに記載の驚き動作生成装置。

【請求項10】

音声信号を受信し、当該音声信号から、その話者の驚き状態を検出し、前記驚き区間信号を生成するための驚き区間信号生成手段をさらに含む、請求項1～請求項9のいずれかに記載の驚き動作生成装置。

【請求項11】

前記音声信号からフォルマントを抽出するフォルマント抽出手段と、

前記フォルマント抽出手段により抽出されたフォルマントに対応して前記ロボットの口唇の開口量を制御するための口唇動作制御手段をさらに含む、請求項10に記載の驚き動作生成装置。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、いわゆるヒューマノイドロボットの制御の改善技術に関し、特に、人間がヒューマノイドロボットを相手に対話しているときのヒューマノイドロボットの動作を従来のものより人に近い自然なものにする技術に関する。

【背景技術】

【0002】

驚き発話は日常会話の場面で頻出し、コミュニケーションを円滑に進める上で感情や態度の表現において重要な役割を果たす。

【0003】

これまで人型ロボットを通して円滑な遠隔コミュニケーションを実現する研究が行われてきたが、人型のヒューマノイドロボット、特に人の姿を持つアンドロイドの場合は、より人らしい振る舞いが要求されることが明らかになっている。特に、驚きのような感情表現において、声は驚いているのに適切な表情又は動作が伴わない場合は、コミュニケーションの相手に違和感が生じる。したがって、驚き発話に伴う自然な動作を生成することは重要である。

【0004】

過去には、笑いに伴うアンドロイドの動作について自然な動作を実現することを目指した技術を開示したもの（後掲の非特許文献1）はあるが、対話において生じる驚きに伴う自然な動作をアンドロイドに行わせるような技術は実現できていない。

【先行技術文献】

【非特許文献】

【0005】

【非特許文献1】 Ishi, C., Funayama, T., Minato, T., Ishiguro, H. (2016).

“Motion generation in android robots during laughing speech,” Proc.

IROS 2016, pp. 3327-3332, Oct., 2016.

【発明の概要】

【発明が解決しようとする課題】

【0006】

既に述べたとおり、アンドロイドの様に人間によく似たロボットの場合、対話中に人間と異なる動作をすると違和感が生じ、かえって不気味な印象を相手に与えてしまう。その

10

20

30

40

50

ような問題を生じさせないためには、対話中のアンドロイドの動作をごく自然なものにする必要がある。特に、驚き動作は、対話を滑らかに進行させる上で非常に重要であるにもかかわらず、そうした問題に着目してアンドロイドの動作を制御する技術はこれまで存在していなかった。

【 0 0 0 7 】

したがって、ヒューマノイドロボットにおいて、驚きに伴う、自然な、人らしい動作を実現するための制御装置を提供することが本発明の1つの目標である。

【 0 0 0 8 】

また、驚きに伴う、自然な、人らしい動作を実現するためにヒューマノイドロボットの各部を協調して動作させる制御装置を提供することも好ましい。

10

【 課題を解決するための手段 】

【 0 0 0 9 】

第1の局面に係る驚き動作生成装置は、所定の時間区間において、驚きに伴う動作を行う様にヒューマノイド型のロボットを制御する。ロボットの顔の表情、頭部の状態、及び上半身の状態は、驚いていない状態である中立状態と、驚いた状態である驚き状態とにそれぞれ制御可能である。この驚き動作生成装置は、上記時間区間の少なくとも開始時刻と終了時刻とを特定する驚き区間信号を受信する様に接続され、開始時刻に、顔の表情を驚き状態に変化させる動作を開始し、終了時刻の後、第1の戻り時間内に、顔の表情を中立状態に戻すよう、ロボットを制御する表情制御手段と、開始時刻に、頭部を驚き状態に変化させる動作を開始し、終了時刻の後、第2の戻り時間内に、頭部の状態を中立状態に戻すよう、ロボットを制御する頭部制御手段と、開始時刻に、上半身を驚き状態に変化させる動作を開始し、終了時刻の後、第3の戻り時間内に、上半身を中立状態に戻すよう、ロボットを制御する上半身制御手段とを含む。表情制御手段、頭部制御手段、及び上半身制御手段のいずれにおいても、第1、第2又は第3の戻り時間が、開始時刻に開始した、驚き状態への変化に要する時間よりも長くなる様に制御される。

20

【 0 0 1 0 】

好ましくは、驚き状態は、第1のレベルの驚き状態と、第1のレベルより高い第2のレベルの驚き状態とを含む。また、驚き区間信号に加えて、驚きのレベルを示す驚きレベル信号がさらに驚き動作生成装置に与えられる。表情制御手段、頭部制御手段、及び上半身制御手段の任意の組み合わせにおいて、驚きレベル信号に応じて驚き状態を第1のレベルに
30

【 0 0 1 1 】

より好ましくは、中立状態から第2の驚き状態へのロボットの各部の変位量は、中立状態から第1の驚き状態へのロボットの各部の変位量より大きい。

【 0 0 1 2 】

さらに好ましくは、表情制御手段は、驚き状態における眉毛の目からの離間量が、中立状態における離間量より大きくなる様に眉毛を制御する。

【 0 0 1 3 】

頭部制御手段は、驚き状態における頭部が、中立状態における位置からロボットにとって後方に移動する様に頭部を制御してもよい。

40

【 0 0 1 4 】

好ましくは、上半身制御手段は、驚き状態における上半身が、中立状態における位置からロボットにとって後方に反る様に上半身を制御する。

【 0 0 1 5 】

より好ましくは、表情制御手段、頭部制御手段、及び上半身制御手段のいずれも、中立状態と驚き状態との間の変化が滑らかになる様にロボットを制御する。

【 0 0 1 6 】

さらに好ましくは、驚き動作生成装置は、ロボットの顔の表情が、驚き状態から中立状態に戻ったことに応答して、ロボットの目を1回閉じさせた後、中立状態に戻す制御を行
50

うことによりロボットに瞬きを行わせるための瞬き制御手段をさらに含む。

【 0 0 1 7 】

驚き動作生成装置は、時間区間において、ロボットが発声すべき発話音声信号のフォルマントを抽出するフォルマント抽出手段と、フォルマント抽出手段により抽出されたフォルマントに対応してロボットの口唇の開口量を制御するための口唇動作制御手段をさらに含んでもよい。

【 0 0 1 8 】

好ましくは、驚き動作生成装置は、音声信号を受信し、当該音声信号から、その話者の驚き状態を検出し、驚き区間信号を生成するための驚き区間信号生成手段をさらに含む。

【 0 0 1 9 】

より好ましくは、驚き動作生成装置は、音声信号からフォルマントを抽出するフォルマント抽出手段と、フォルマント抽出手段により抽出されたフォルマントに対応してロボットの口唇の開口量を制御するための口唇動作制御手段をさらに含む。

【 図面の簡単な説明】

【 0 0 2 0 】

【 図 1 】 分析データにおける、形態素ごとの驚き発話の分布を示す図である。

【 図 2 】 分析データ中の驚き発話の驚きの度合いによる分布を表形式で示す図である。

【 図 3 】 分析データ中の驚き発話に伴う動作タイプの分布を示す図である。

【 図 4 】 分布データ中の各驚き表現における動作の出現度を示す図である。

【 図 5 】 分布データ中の驚き表現に伴う各動作タイプに対する動作の出現度を示す図である。

【 図 6 】 分布データ中の各形態素に関する動作タイプの分布を示す図である。

【 図 7 】 眉毛の引き上げ時間及び戻り動作の持続時間の平均及び標準偏差を驚きのレベル別に示す図である。

【 図 8 】 上半身戻り動作における開始及び戻りに要する時間の平均と標準偏差とをレベル 1 及びレベル 2 の双方について示した図である。

【 図 9 】 各動作タイプについて、動作と驚き発話との間の時間差の分布を示す図である。

【 図 1 0 】 アンドロイドに設けられたアクチュエータの配置を示す図である。

【 図 1 1 】 アンドロイドの中立状態の表情、レベル 1 の驚き表情、及びレベル 2 の驚き表情を示す図である。

【 図 1 2 】 アンドロイドの制御装置の構成を示すブロック図である。

【 図 1 3 】 図 1 2 に示す制御装置を実現するコンピュータのハードウェア構成を示すブロック図である。

【 発明を実施するための形態】

【 0 0 2 1 】

以下の説明及び図面では、同一の部品には同一の参照番号を付してある。したがって、それらについての詳細な説明は繰返さない。

【 0 0 2 2 】

本実施の形態は、驚き発話に焦点を当て、対面対話データを分析し、驚き表現の度合い又は種類と、それに伴う表情及び身体動作との関連性を分析し、その結果に基づいて人らしい驚き動作を実現した。

【 0 0 2 3 】

[分析データ]

〈 マルチモーダル対話音声データベース 〉

分析には、出願人が収集したマルチモーダル対話音声データベースを用いた。このデータベースはさまざまな年代の話者による、音声、頭部のモーションキャプチャデータおよびビデオデータを含む。各対話は 1 0 分程度で、自由会話である。データベースは発話区間の情報と書き起こしを含み、驚きが表現された発話にはラベルが付与されている。

【 0 0 2 4 】

データベースには、フレーズ（アクセント句）毎に談話機能ラベルが付与され、感動詞

10

20

30

40

50

には、相槌、肯定、聞き返し、意外、驚き、感心、気づきなどのパラ言語ラベルが付与されている。書き起こしには感嘆詞「！」が付いているものもあり、感動詞以外のフレーズで驚きを表現したものも含まれている。そこで、「！」を含むフレーズを抽出し、明らかに驚き以外のものは除外して、驚きラベルが付与された感動詞と合わせて分析データとして用いた。分析には28名の成人話者（男性12名、女性16名）のデータを用いた。このデータからおよそ636個の驚き発話が抽出された。

【0025】

図1に形態素ごとの驚き発話の分布を示す。感動詞「え」が最も出現度が多く(40%)、感動詞「あ」(20%)と「へ」(11%)が次いで多く出現した。

【0026】

10

〈驚きの度合と種類のラベルデータ〉

各驚き発話において、驚きの度合を4段階（驚いていない、少し驚いている、驚いている、すごく驚いている）で付与するようラベル付与者に指示した。その際に、会話の文脈を考慮するため、発話の前後5秒の区間も含め、対話の相手の声も聴ける様にした。

【0027】

驚き発話の中には、感情的・自発的なものと意図的に表現したもののが混合している。そこで、自発的か意図的かに関わらず、驚きの表現の度合を判断するようラベル付与者に指示した。驚きの表現の種類は以下の項目で付与するよう指示した。

【0028】

20

- 1 感情的・自発的な反応
- 2 社会的・意図的な反応
- 3 対話の中に過去の驚き発話を再現・引用

意図的な驚き表現は、対話インタラクションを円滑にするため、自然に発生することが多い。しかし、驚き表現が自発的か意図的かを第三者が判断するのは、ラベル付与者の主観によるため、判断結果にばらつきが多く生じると予想される。

【0029】

日本語母語話者4名が、各発話において驚きの度合および驚きの種類のラベルを付与した。驚きの度合は、平均を取り、0～3の指標で表現した。各度合の出現数は図2に示すとおり、レベル0：63個、1：361個、2：187個、3：25個となった。日常会話で最も多く出現するのは、驚き表現の度合が低いもの（レベル1）である。

30

【0030】

驚いていないと判断された63個の発話（レベル0）は除外し、残りの573個の発話を分析対象とした。驚き表現の種類に関しては、出現数が{自発的：343個(60%)、意図的：187個(33%)、再現：43個(7.5%)}となった。

【0031】

〈驚き発話に伴う動作のラベルデータ〉

驚き発話に伴う表情や身体動作に関連するラベルとして、以下の項目を使用した。

【0032】

瞼：{閉じている、普通に開いている、大きく見開いている}

眉毛：{上がっている、少し上がっている、上がっていない、顰めている}

40

頭部：{動いていない、上、下、左か右、上昇下降、傾げ、傾き、その他}

胴体：{動いていない、前、後ろ、上、下、左か右、傾げ、回る、その他}

各驚き発話区間に対し、研究補助員1名がビデオを見ながら表情や動作に関連するラベルを付与した。複数の動作が含まれる場合は、複数項目の選択を可能とした。

【0033】

図3に驚き発話で観測された動作の分布を示す。図3を参照して、眉毛を上げる動作（目を見開く動作に伴うことが多い）が最も多く、次いで頭部を上げるまたは上げてすぐ下げる動作、上半身を後ろに反る動作、及び傾き動作が多く出現した。動作が全く伴わない驚き発話は30%で観測された。これは日常会話で驚きを表現する度に表情や動作が必ずしも伴わないことを示している。また、動作の出現度は、驚き表現の度合にも依存すると

50

考えられる。

【 0 0 3 4 】

〔 分析結果 〕

図4に、驚き表現の種類別で、驚きの度合における動作の出現度の分布を示す。全体的に驚き表現の度合が上がるにつれて、動作の出現度も上がることが分かる(「全体」)。この傾向は、感情的・自発的な驚き発話でより顕著に現れる。一方で、社会的・意図的な驚き発話および驚きの引用発話では、驚きの度合が高くても動作が必ずしも伴わないことが分かる。

【 0 0 3 5 】

図5に、出現度の高かった動作(眉毛、上半身、頭部)において、驚きの度合における動作の出現度の分布を示す。この結果より、眉毛を上げる動作は、驚きの度合が中および高(レベル2及び3)の場合に高くなるのに対し、上半身を後ろに反る動作は、驚きの度合が高(レベル3)の場合のみ高くなる。

10

【 0 0 3 6 】

図6に、出現度が最も多かった感動詞「え」、「あ」および「へ」における各動作の出現度の分布を示す。感動詞によって、驚き発話に伴う動作の傾向も変わることが分かる。「え」の場合は、眉毛を上げる動作が最も多く、「あ」の場合は頭を上げる動作が多かった。「へ」の場合は、頷き動作が最も多く、驚きと感心や同情の表現が伴うことが原因と考えられる。

【 0 0 3 7 】

20

以上の分析の結果、驚き表現の種類、度合、及び形態素によっても、生成すべき動作が異なることが明らかとなった。

【 0 0 3 8 】

さらに、驚き表現において重要なのは、各動作について、発話と自然な形で同期するように、動作の開始及び戻りのタイミングをコントロールすることである。そのための驚き発話の分析結果について以下に説明する。

【 0 0 3 9 】

眉毛及び胴部の開始及び戻りに要する時間を、「え」及び「あ」という感動詞について測定した。結果を図7に示す。図7には、眉毛の引き上げと戻りの時間について、2つのレベルについて測定した結果の平均と標準偏差とを示す。レベル1の開始に要する時間(開始時間)は「開始1」、戻りは「戻り1」で示し、レベル2の開始時間を「開始2」、戻り時間を「戻り2」により示す。

30

【 0 0 4 0 】

図7から明らかな様に、いずれのレベルでも開始時間の方が戻りに要する時間より短い。開始時間の平均は200から300ミリ秒、戻り時間の平均は400から500ミリ秒である。レベル2はレベル1より多少時間が長くなっているが、動く距離が長いためである。レベル2における戻り時間の標準偏差はレベル1の戻り時間の標準偏差よりはるかに大きい、これは眉毛が通常的位置に戻るのに1秒から2秒を要するためである。アンドロイドによる驚き動作でもこれと同様の制御をすることが望ましい。これらの制御をするための時間を、図7に示すような標準偏差に基づいて変化させる様にしてもよい。これは

40

【 0 0 4 1 】

図8は、上半身反り動作における開始及び戻りに要する時間の平均と標準偏差とをレベル1及びレベル2の双方について示したものである。上半身については、開始時間及び戻り時間がレベル1ではいずれも0.8秒程度、レベル2ではそれぞれ1.2秒及び1.5秒であった。アンドロイドによる驚き動作でもこれと同様の制御をすることが望ましい。この時間をばらつかせる様にしてもよいことは眉毛の引き上げの場合と同様である。

【 0 0 4 2 】

図9は各動作タイプについて、動作と驚き発話との間の時間差を示す。「開始」は動作の開始時刻と発話の開始時刻との間の差を示す。「終了」は動作の終了時刻と驚き発話の

50

終了時刻との間の差を示す。この結果から、眉毛、頭部、及び上半身の動作の開始時刻と、驚き発話の開始時刻との差はほぼ-0.1秒~0.1秒の間であることが分かった。即ち、通常、驚きに伴う動作は驚き発話とほぼ同期している。これに対し、終了時刻の差については、よりばらついているが大部分がプラスであることが分かった。すなわち、驚き動作の後に体が通常の中立状態の位置に戻るのは、驚き発話が終わってからということである。

【0043】

〔アンドロイドロボットにおける動作生成〕

本実施の形態では、驚き発話の持続時間が所与のものであるとして、種々の驚き度合いについて種々の驚き動作を制御する場合の効果について検討した。分析の結果、発明者は、以下の4つの要件を考慮した動作生成方法を提案する。すなわち、表情制御（瞼の開大を伴う眉毛の引き上げ）、頭部動作制御（頭部のピッチ方向）、及び上半身動作制御（トルソのピッチ方向）である。

10

【0044】

＜アンドロイドのアクチュエータと制御方法＞

本実施の形態で制御対象となるのは女性のアンドロイドであり、このアンドロイドを用いて驚き発話に伴う動作を評価した。アンドロイドは、遠隔の話者から送られてくる音声信号に応じて、アンドロイドの前にいる相手と対話することが想定されている。アンドロイドの発声する音声は、遠隔の話者から送られてくる音声を再生したものでもよい、別の音声に変換したものでもよい。

20

【0045】

図10に、このアンドロイドの上体の動作を制御するアクチュエータの配置を示す。図に示す様に、このアンドロイドの上半身に設けられたアクチュエータは全部で19個である。図10にはアクチュエータ1~11、13~19を示してある。アクチュエータ12は舌の動きを制御するものであり、図10には示していない。またアクチュエータ17は肩の動きのためのものであり、本実施の形態による驚き動作制御には関与しない。すなわち、これらのうち、アクチュエータ1~16、18及び19が驚き動作における制御の対象となる。

【0046】

表情制御にはアクチュエータ1~13が関与し、頭部動作制御にはアクチュエータ14~16が関与し、上半身制御にはアクチュエータ18及び19が関与する。すなわち、表情制御の自由度は13、頭部動作の自由度は3、上半身制御の自由度は2である。

30

【0047】

驚き動作に関与するアクチュエータは以下のとおりである。

【0048】

上瞼制御（アクチュエータ1）

下瞼制御（アクチュエータ5）

眉毛引き上げ制御（アクチュエータ6）

口角引き上げ制御（アクチュエータ8。ほほも同時に引き上げられる。）

口角伸長制御（アクチュエータ10）

40

顎部引き下げ制御（アクチュエータ13）

頭部ピッチ制御（アクチュエータ15）

上半身ピッチ制御（アクチュエータ18）

全てのアクチュエータについて、コマンド値は0~255の範囲で与えられる。なお、以下の説明では、アクチュエータの状態（位置）については、そのコマンド値を用いて表し、「アクチュエータ値」と呼ぶことがある。

【0049】

- 眉毛及び瞼の動作制御 -

驚き動作における顔の表情について、眉毛の引き上げと瞼の開大は、驚きの2つのレベルにしたがって互いに協働するよう制御される。眉毛のアクチュエータ6の位置はレベル

50

1 については1 2 7、レベル2 については2 5 5とし、上瞼及び下瞼のアクチュエータ1 及び5 についての目標値はレベル1 についてはそれぞれ8 0 及び6 0、レベル2 についてはそれぞれ4 0 及び3 0とする。中立状態の位置(レベル0 に対応)における顔表情については、アクチュエータ6、1、5 の位置はそれぞれ0、9 0 及び8 0 である。これらの値は、アンドロイドロボットの顔の表情を見ながら手動により設定した。この設定により、レベル1 については少し驚いた表情、レベル2 については明らかな驚きの表情が得られた。

【0 0 5 0】

図1 1 に、これらに対応するアンドロイドの顔の写真を示す。図1 1 (A) は中立状態、(B) はレベル1 の驚きの表情、(C) はレベル2 の驚きの表情である。

10

【0 0 5 1】

上記した分析結果にしたがい、瞼及び眉毛のアクチュエータ1、5 及び6 に対しては、驚き発話の開始と同時に、各レベルに応じた値を送信し、驚き発話が終了してから0.5 秒経過するまでに、滑らかに中立状態の位置に戻るようコサイン関数によりコマンドを生成する。また、上記分析に伴い、顔の表情が中立状態に戻ったときに瞬きが発生することが多いということが判明したので、本実施の形態でも、顔の表情を中立状態に戻したときに目を短時間(1 0 0 ミリ秒)閉じて中立状態に戻すことで瞬きを実現した。具体的には、アクチュエータ1 及び5 にコマンド2 5 5 を送って1 0 0 ミリ秒の間目を閉じさせ、その後それぞれコマンド9 0 及び8 0 を送って中立状態の位置に戻す様にした。

20

3. 1. 2 上半身動作制御

上半身動作制御に関し、本実施の形態では以下の式(1) に示される様に、コサイン関数にしたがって上半身を後方に動かす。式(1) において「act[18][t]」は、アクチュエータ1 8 に対して時刻t 秒に送信されるコマンドを表す。

【0 0 5 2】

【数1】

$$act[18][t] = 32 + (upbody_{target} - 32) * \frac{1 - \cos\left(\pi \frac{t - t_{start}}{T_{onset}}\right)}{2} \quad (1)$$

式(1) において t_{start} は驚き動作の開始時刻、すなわち驚き発話の開始時刻であり、 T_{onset} は驚き動作の開始動作の持続時間である。また $upbody_{target}$ は驚き動作に伴う上半身の反り量の最大値を示す。式中の「3 2」という値はアンドロイドの中立状態の姿勢におけるアクチュエータ1 8 の位置である。式(1) にしたがって驚き動作を行わせることにより、所与の時間 T_{onset} にアンドロイドの上半身の反り量を滑らかに所与の最大値 $upbody_{target}$ に到達させることができる。

30

【0 0 5 3】

$upbody_{target}$ の値は、本実施の形態では、レベル1 については1 6、レベル2 については0 である。ただし、本実施の形態では、アクチュエータ値が小さい程上半身が後ろ側に反る設定である。もちろん、これらの値は設計により、使用するアクチュエータにより、それぞれ変えることができる。

40

【0 0 5 4】

驚き発話の終了時刻から、上体を次の式(2) に示す様にコサイン関数にしたがって中立状態の位置に戻す。

【0 0 5 5】

【数2】

$$act[18][t] = 32 + (upbody_{end} - 32) * \frac{1 + \cos\left(\pi \frac{t - t_{end}}{T_{offset}}\right)}{2} \quad (2)$$

式(2) において $upbody_{end}$ 及び t_{end} はそれぞれ、驚き発話の終了時におけるアクチュエ

50

ータ18の値及び時刻である。したがって、驚き発話の発話区間が T_{onset} よりも短いと、上半身の反り量は最大値まで達しない。

【0056】

上記した分析の結果、本実施の形態における驚き動作において、動作の開始から最大値に到達するまでの持続時間は0.8秒に設定した。驚き発話の終了時から中立状態の位置に戻るまでの時間 T_{offset} は1.5秒に設定した。

【0057】

-頭部動作制御-

頭部動作に関し、本実施の形態では、頭部のピッチを音声ピッチ（基本周波数 F_0 ）により制御する方法を採用した。実際の人間がこのような頭部動作をしているわけではないが、笑いに関する同様の研究から、このような制御をしても自然な頭部動作を実現できることが判明しているため、本実施の形態でも同様の制御を行うこととした。通常、驚いたときには、人間は高い F_0 の音声を発生して頭を上に向かって顔を上に向け、又は上に向けた後に下に向ける動作をすることが多いためである。本実施の形態では、この F_0 の値を頭部ピッチアクチュエータ15へのコマンド値に変換するために、以下の式(3)を用いる。

【0058】

【数3】

$$act[15][t] = 128 + (F_0[t] - center_F_0) * F_0_scale \quad (3)$$

式(3)において、 $center_F_0$ は発話者の平均 F_0 の値（男性は約120Hz、女性は約240Hz）であって、半音単位に変換されたものであり、 F_0 は現在の F_0 の値（半音値）であり、 F_0_scale は F_0 の変化を頭部のピッチ動作にマッピングするためのスケールファクタである。本実施の形態では、 F_0_scale は音声ピッチの1半音の変化が頭部のピッチ方向の回転の1度にほぼ対応する様に設定した。

【0059】

予備実験から、驚き発話の際に上半身が後方にそったときの顔が上を向くと不自然に感じられることが分かった。また、観察により、通常の話者ではそうした場合でも顔を対話の相手方に向けていることが分かった。そのため、頭部ピッチアクチュエータ15について、上半身の後方への反り動作の際に、逆方向の動きを加える制御を行うことにした。そのため、コマンド値は以下の式(4)により計算している。

【0060】

【数4】

$$act[15][t] = act[15][t] - (32 - act[18][t]), \quad (act[18][t] < 32 \text{ のとき}) \quad (4)$$

-口唇動作制御-

有声で発話されない驚き表現では、顎部を下げるべきである。しかし、有声で発話される驚き表現では、口唇の動きと、それに続く顎部の動きは、ともに発話内容と同期すべきである。本実施の形態では、口唇動作は特開2012-173389号公報において提案された、フォルマントに基づく口唇動作制御方法を採用した。この様にすることで、種々の母音音質を持つ有声の驚き発話部分（「えー」、「あー」等）における適切な口唇形状を生成できる。この方法は母音のフォルマントを元に行っているためである。顎部アクチュエータ（アクチュエータ13）はこの様にして推定された口唇の高さを用いて制御する。

【0061】

〔構成〕

以上、制御の細部について説明した本実施の形態に係る驚き動作の生成装置は以下のような構成を有する。図12を参照して、この実施の形態に係る驚き動作生成装置200は音声信号202を受けて、発話の内容、韻律、及び声質に基づいて驚き区間を検出し、その開始と終了とを少なくとも示す驚き区間信号を生成し出力するための驚き区間信号生成部210と、音声信号202からフォルマントを抽出しフォルマントの値を示す信号を出

10

30

40

50

力するフォルマント抽出部212と、フォルマント抽出部212の出力する信号に基づいて驚き区間中の口唇動作のためのアクチュエータ群（図示せず）を制御するための口唇動作制御部214とを含む。遠隔の話者の音声に基づいて、その話者の驚き状態を検出し、その驚き状態に応じ、実際の人間に近い自然な動きでアンドロイドの表情、頭部、及び上半身を制御する。

【0062】

本実施の形態では、アンドロイドの状態（眉毛の位置及び瞼の開大量、頭部の位置及び上半身の状態）は、中立状態、第1のレベルの驚き状態、第2のレベルの驚き状態に制御可能である。それら状態は、各部を動作させるアクチュエータへのコマンド値をそれぞれ所定の値に制御することにより実現できる。

10

【0063】

なお、以下の説明から明らかな様に、第2のレベルの驚き状態における各部の、中立状態における位置からの離間量は、第1のレベルの驚き状態における離間量よりも大きい。また、本実施の形態では、各部を第1又は第2のレベルの驚き状態から中立状態に戻すのに要する時間が、中立状態から第1又は第2のレベルの驚き状態に変化させるのに要する時間よりも長くなる様にアクチュエータを制御する。以下の実施の形態では表情、頭部、及び上半身のいずれも、第1及び第2のレベルの驚き状態とで互いに異なる動作をしている。しかし、これらがいずれもそのような動作をする必要はない。表情、頭部、及び上半身の任意の組み合わせについて、第1及び第2のレベルの驚き状態によって制御を異ならせる様にし、残りの部分については同じ制御を行う様にしてもよい。

20

【0064】

フォルマント抽出部212としては、上述の通り特開2012-173389号公報に開示された口唇動作パラメータ生成装置と同様の構成を採用すれば良い。

【0065】

驚き区間信号生成部210は、音声信号202に対して発話内容の音声認識を行い、認識結果である発話内容のテキストを出力する音声認識装置260と、音声信号202から音声の韻律及び声質特徴を抽出するための韻律・声質特徴抽出部262と、音声認識装置260から出力される発話内容のテキストと、韻律・声質特徴抽出部262が出力する韻律・声質特徴とを用いて発話中の驚き区間の開始及び終了位置、並びに驚きのレベル（1、2）を検出し、驚き区間信号を出力するための驚き区間検出部264とを含む。韻律・声質特徴抽出部262が抽出する特徴は、F0、規則性、及び力み等である。驚き区間検出部264としては、例えば特開2010-217502号公報に開示された発話意図情報検出の方法を採用すればよい。

30

【0066】

驚き動作生成装置200はさらに、驚き区間検出部264が出力する驚き区間検出信号に応答して、アンドロイドの表情を制御する信号を生成するための表情制御部216を含む。

【0067】

表情制御部216は、驚き区間信号に応答して、その開始時刻及び終了時刻をトリガーにアンドロイドの眉毛の動作をさせるためのアクチュエータ信号を生成し、眉毛に関するアクチュエータ群224（アクチュエータ6）に出力するための眉毛引き上げ動作制御部280と、驚き区間信号に応答して、その開始時刻及び終了時刻をトリガーに、瞼の開大動作をさせるためのアクチュエータ信号を生成し、瞼の動作を制御するアクチュエータ群226（アクチュエータ1及び5）に出力するための瞼開大動作制御部282とを含む。本実施の形態では、第1のレベルの驚き状態における眉毛の引き上げ量よりも、第2のレベルの驚き状態における引き上げ量の方が大きくなる様に制御を行う。瞼の開大量も同様である。

40

【0068】

驚き動作生成装置200はさらに、驚き区間の終了時、驚き動作から中立状態の位置にアンドロイドの姿勢等が戻ったときにアンドロイドに瞬きをさせるためのアクチュエータ

50

信号を生成し、瞼の動作を制御するアクチュエータ群226に出力する瞬き制御部218と、驚き区間検出部264が出力する驚き区間信号に応答して、韻律・声質特徴抽出部262が出力する韻律・声質特徴と驚き区間信号により示される驚き区間の開始及び終了に基づいてそれぞれ驚き動作の開始及び終了時の頭部の動作を制御するコマンドを生成し、頭部動作に関するアクチュエータ群228（アクチュエータ15）に対して出力する頭部動作制御部220と、驚き区間検出部264が出力する驚き区間信号に応答して、韻律・声質特徴抽出部262が出力する韻律・声質特徴と驚き区間信号により示される驚き区間の開始及び終了に基づいてそれぞれ驚き動作の開始及び終了時の上半身の動作を制御するコマンドを生成し、腰部アクチュエータ群230に出力する上半身動作制御部222とを含む。

10

【0069】

本実施の形態では、頭部ピッチの変化量（動き角度）及び上半身の反り量の何れも、第1のレベルの驚き状態における値より第2のレベルの驚き状態における方が大きくなる様に制御される。なお、驚きの開始時と終了時とにおける各部の動作はできるだけ滑らかにすることが望ましい。そのために、例えば式（1）及び（2）に示すような式で位置を制御することが望ましい。

【0070】

〔動作〕

遠隔から送信されてきた話者の音声信号202が与えられると、驚き動作生成装置200は以下の様に動作する。音声認識装置260は音声信号202に対する音声認識を行い、発話内容のテキストを驚き区間検出部264に出力する。韻律・声質特徴抽出部262は、音声信号202から発話音声の韻律及び声質特徴を抽出し、驚き区間検出部264に与える。驚き区間検出部264は、発話内容のテキスト及び韻律及び声質特徴に基づいて遠隔の話者の驚き状態を反映した驚き区間と驚きレベルとを検出し、少なくともその開始時刻及び終了時刻と驚きレベルとを特定する驚き区間信号を生成して、眉毛引き上げ動作制御部280、瞼開大動作制御部282、瞬き制御部218、頭部動作制御部220及び上半身動作制御部222に与える。

20

【0071】

眉毛引き上げ動作制御部280及び瞼開大動作制御部282は、驚き発話の開始と同時に、各驚きレベルに応じた値をアクチュエータ群224及び226にそれぞれ送信し、驚き発話が終了してから0.5秒経過するまでに、滑らかに中立状態の位置に戻るようコサイン関数によりコマンドを生成しアクチュエータ群224及び226にそれぞれ送信する。

30

【0072】

瞬き制御部218は、驚き区間の終了時、驚き動作から中立状態の位置にアンドロイドの姿勢等が戻ったときにアンドロイドに瞬きをさせるためのアクチュエータ信号を生成し、アクチュエータ群226に与える。具体的には、アクチュエータ1及び5にコマンド255を送って100ミリ秒の間、瞼を閉じさせ、その後にそれぞれコマンド90及び80を送って瞼を中立状態の位置に戻すことによりアンドロイドに瞬きを行わせる。

【0073】

頭部動作制御部220は、驚き区間信号に応答して、韻律・声質特徴抽出部262からの韻律・声質特徴を用いて頭部の動作を制御する。具体的には、頭部動作制御部220は、驚き区間の開始時には、韻律・声質特徴抽出部262から与えられるF0を式（3）及び式（4）を用いて頭部動作のためのコマンド値を算出し、アクチュエータ15に対して出力する。ただし式（4）による変換を行うのは、上半身動作制御部222から上半身の動作のためのアクチュエータ群230（アクチュエータ18）に与えられるコマンドが32より小さいときだけである。

40

【0074】

上半身動作制御部222は、驚き区間信号に応答し、驚き区間の開始時には式（1）にしたがって驚きレベルに応じたコマンドを生成しアクチュエータ群230に与える。また

50

上半身動作制御部222は、驚き区間の終了時には式(2)にしたがって驚きレベルに応じたコマンドを生成しアクチュエータ群230に与えて上半身を中立状態の位置に戻す。
【0075】

一方、フォルマント抽出部212は驚き区間信号生成部210からフォルマントを抽出し口唇動作制御部214に与える。口唇動作制御部214はフォルマントに基づく口唇動作制御方法にしたがい、種々の母音音質を持つ有声の驚き発話部分に応じた適切な口唇形状を生成するためのコマンドを生成し、口唇動作を行うアクチュエータ13に与えて口唇及び顎部を動作させる。口唇動作制御部214は驚き信号とは無関係に動作するが、この様にフォルマントに基づいて口唇動作を制御しても、音声の声質等に応じて驚き区間でも適切な口唇動作を実現できる。

10

【0076】

[コンピュータによる実現]

図13は、アンドロイドを制御する装置の典型例の構成を示す。この実施の形態では、制御装置630はコンピュータ640からなり、このコンピュータ640をアンドロイドの各部と接続することでアンドロイドの動作を制御する。図13を参照して、制御装置630は、メモリポート652及び入出力インターフェイス(入出力I/F)650を有するコンピュータ640と、いずれもコンピュータ640に接続されたキーボード646と、マウス648と、モニタ642とを含む。コンピュータ640は、入出力I/F650を介してアンドロイドの各部と接続されている。

【0077】

20

コンピュータ640はさらに、CPU(中央処理装置)656と、CPU656、メモリポート652及び入出力I/F650に接続されたバス666と、起動プログラム等を記憶する読出専用メモリ(ROM)658と、バス666に接続され、上記領き生成装置150の各部の機能を実現するプログラム命令、システムプログラム及び作業データ等をプログラムの実行時に記憶するランダムアクセスメモリ(RAM)660と、ハードディスク654を含む。コンピュータ640はさらに、他端末との通信を可能とするネットワーク668への接続を提供するネットワークインターフェイス(I/F)644を含む。

【0078】

制御装置630を上記した実施の形態に係るアンドロイドの制御装置として機能させるためのコンピュータプログラムは、メモリポート652に装着されるリムーバブルメモリ664、又は入出力I/F650に接続される図示しない外部記憶装置に記憶され、さらにハードディスク654に転送される。又は、プログラムはネットワーク668を通じてコンピュータ640に送信されハードディスク654に記憶されてもよい。プログラムは実行の際にRAM660にロードされる。図示しない外部記憶装置から、リムーバブルメモリ664から又はネットワーク668を介して、直接にRAM660にプログラムをロードしてもよい。

30

【0079】

このプログラムは、コンピュータ640を、上記実施の形態に係るアンドロイドの制御装置の各機能部として機能させるための複数の命令からなる命令列を含む。コンピュータ640にこの動作を行わせるのに必要な基本的機能のいくつかはコンピュータ640上で動作するオペレーティングシステム若しくはサードパーティのプログラム又はコンピュータ640にインストールされる、ダイナミックリンク可能な各種プログラミングツールキット又はプログラムライブラリにより提供される。したがって、このプログラム自体はこの実施の形態のシステム、装置及び方法を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令のうち、所望の結果が得られる様に制御されたやり方で適切な機能又はプログラミングツールキット又はプログラムライブラリ内の適切なプログラムを実行時に動的に呼出すことにより、上記したシステム、装置又は方法としての機能を実現する命令のみを含んでいればよい。もちろん、独立したプログラムのみで必要な機能を全て提供してもよい。

40

【0080】

50

上記実施の形態では、図12に示す驚き区間検出部264にルールベースのものを用いている。しかし本発明はそのような実施の形態には限定されない。ルールベースの検出部に代えて、SVM、ナイーブベイズ、ロジスティック回帰、DNN、RNN、CNN等からなる、機械学習による判定器を用いても良い。

【0081】

さらに、上記実施の形態では、驚き区間信号生成部210による驚き区間の検出を行っている。これは、遠隔の話者の音声に応じて、その話者の驚き状態をアンドロイドに反映させる必要があるためである。本発明はそのような実施の形態には限定されない。例えば、アンドロイドを所定のシナリオにしたがって動作させるような場合には、驚き区間は既知となる。したがって、驚き区間信号生成部210は不要であり、シナリオにしたがって、驚き動作を行うべき区間を示す信号を眉毛引き上げ動作制御部280、瞼開大動作制御部282、瞬き制御部218、頭部動作制御部220及び上半身動作制御部222に与える様にすればよい。この場合には、フォルマント抽出部212も不要で、合成出力すべき音声のターゲットからフォルマントを求めればよい。

10

【0082】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内での全ての変更を含む。

【符号の説明】

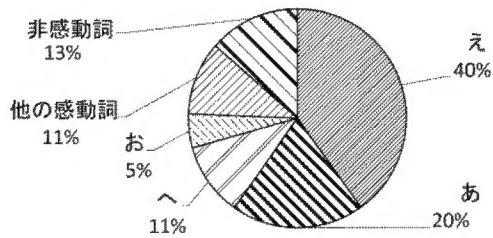
20

【0083】

200 驚き動作生成装置
202 音声信号
210 驚き区間信号生成部
212 フォルマント抽出部
214 口唇動作制御部
216 表情制御部
218 瞬き制御部
220 頭部動作制御部
222 上半身動作制御部
224、226、228、230 アクチュエータ群
260 音声認識装置
262 韻律・声質特徴抽出部
264 驚き区間検出部
280 眉毛引き上げ動作制御部
282 瞼開大動作制御部
630 制御装置

30

【 図 1 】

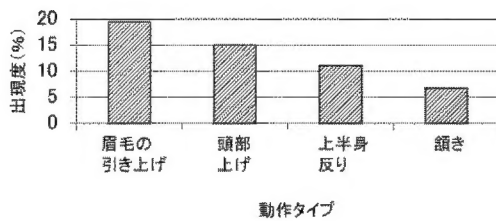


【 図 2 】

知覚上の驚き度合い	0	1	2	3
発話数	63	361	187	25

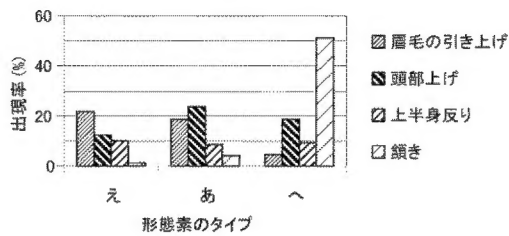
【 図 3 】

驚き表情中の動作タイプの分布



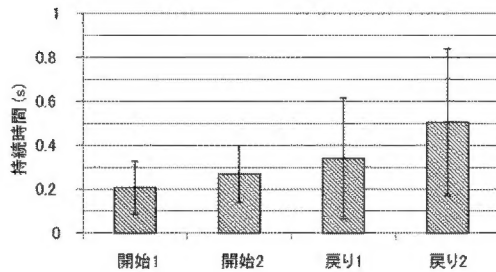
【 図 6 】

各形態素に関する動作タイプの分布



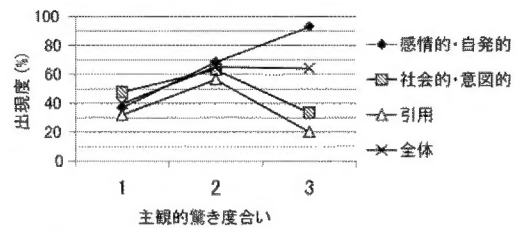
【 図 7 】

眉毛引き上げの開始及び戻り時間



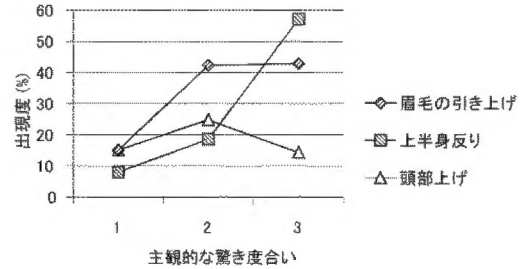
【 図 4 】

各驚き表現における動作の出現度



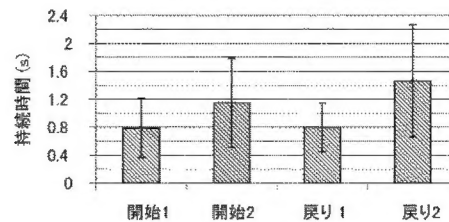
【 図 5 】

各動作タイプに対する動作の出現度



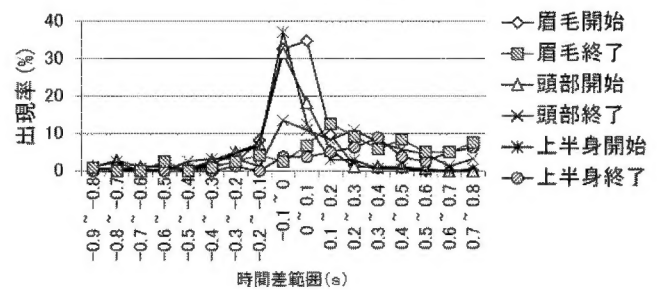
【 図 8 】

上半身反り動作の開始及び戻り時間

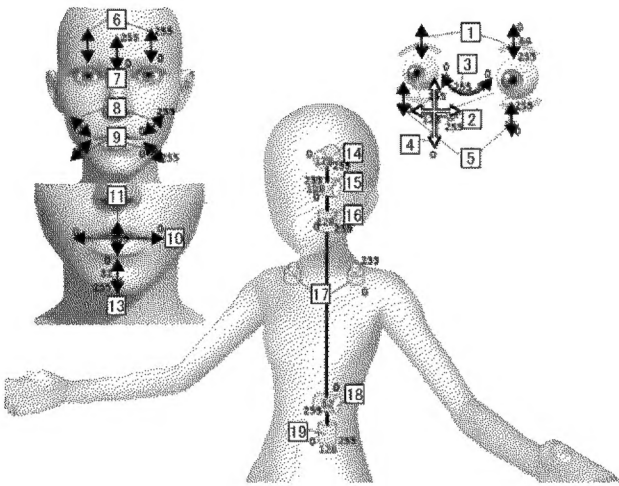


【 図 9 】

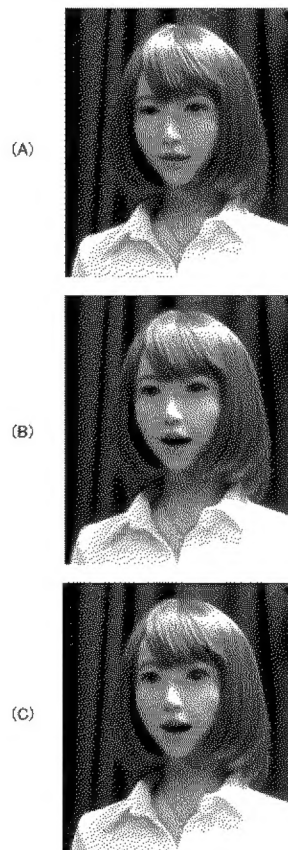
動作と発話間の時間差分布



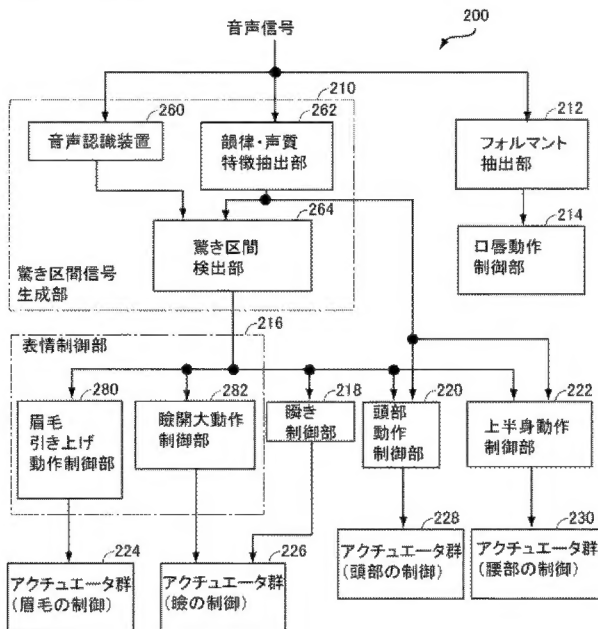
【図10】



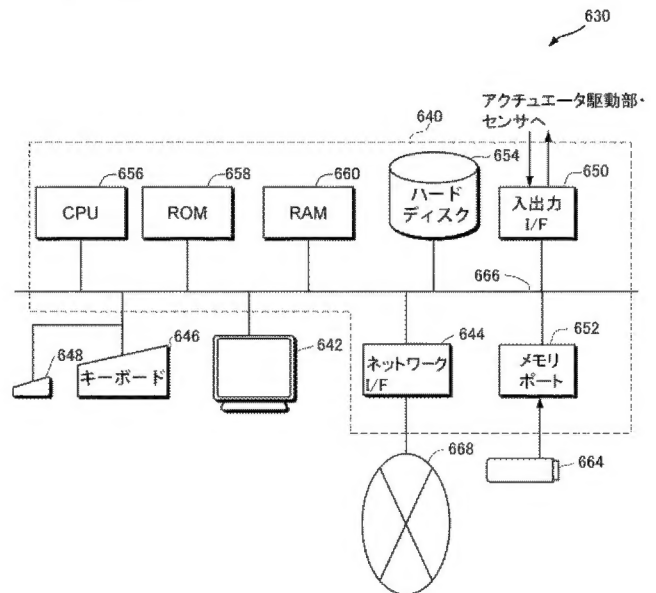
【図11】



【図12】



【図13】



フロント ページの続き

(51)Int.Cl.			F I	テーマコード (参考)		
G 1 0 L	13/00	(2006.01)	G 1 0 L	13/00	1 0 0 A	
G 1 0 L	15/00	(2013.01)	G 1 0 L	15/00	2 0 0 H	
A 6 3 H	13/04	(2006.01)	A 6 3 H	13/04	J	
A 6 3 H	11/00	(2006.01)	A 6 3 H	11/00	Z	